

Title: "STAT 601 – Homework 7"

Author: "Julius Hai"

Date: "10/14/2025"

1a. What are the appropriate assumptions?

```
'data.frame': 25 obs. of 2 variables:
 $ Percent_Sand : Factor w/ 5 levels "10%","15%","20%",...: 1 1 1 1 1 2 2
 2 2 2 ...
 $ Compression_Resistance: num 3100 3200 3150 3250 3180 3300 3350 3320 3400 3370
 ...

Shapiro-wilk normality test

data: residuals(anova_model)
W = 0.98894, p-value = 0.9921

Levene's Test for Homogeneity of Variance (center = median)
Df F value Pr(>F)
group 4 1.3657 0.2811
20
```

1b. What is the appropriate test for the problem?

```
Percent_Sand Df Sum Sq Mean Sq F value Pr(>F)
Residuals 20 72.5 3.62
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Tables of means
Grand mean
39.93334

Percent_Sand
Percent_Sand
5 10 15 20 25
30.39 34.91 40.62 45.22 48.53
```

1c. State the Hypothesis test contains the correct mathematical/statistical notation (Greek letters, subscripts, and symbols, etc.)

```
MODEL:
Y_ij = μ + τ_i + ε_ij, with ε_ij ~ N(0, σ^2)
i = 1,...,5 (sand levels), j = 1,...,n_i

HYPOTHESES:
H0: μ1 = μ2 = μ3 = μ4 = μ5 (all group means equal)
Equivalently: τ1 = τ2 = τ3 = τ4 = τ5 = 0
Ha: ∃ i ≠ j such that μ_i ≠ μ_j (at least one mean differs)

ANOVA TABLE (F-test corresponding to H0 vs Ha):
Df Sum Sq Mean Sq F value Pr(>F)
Percent_Sand 4 1093.6 273.40 75.44 8.94e-12 ***
Residuals 20 72.5 3.62
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Test statistic: F = 75.4382, p-value = 8.937e-12, alpha = 0.05
Decision: Reject H0 → At least one mean differs among sand levels.
```

1d. Conduct the test using R, show the snippet of the results the results

```

      Df Sum Sq Mean Sq F value    Pr(>F)
Percent_Sand  4 264304    66076   52.44 2.53e-10 ***
Residuals    20  25200     1260
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Tables of means
Grand mean

3337.2

Percent_Sand
Percent_Sand
10% 15% 20% 25% 30%
3176 3348 3484 3390 3288

=====
ANOVA Test Result:
F = 52.441, p-value = 0.0000
Significance Level (alpha) = 0.05
Conclusion: Reject H0 → Sand content significantly affects compression resistance.
=====

```

1e. Interpret the results by stating what the alpha level, F test result, and interpret the p value and your conclusion about the means.

```

      Df Sum Sq Mean Sq F value    Pr(>F)
Percent_Sand  4 264304    66076   52.44 2.53e-10 ***
Residuals    20  25200     1260
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
=====
ANOVA INTERPRETATION SUMMARY
=====
Significance Level (α): 0.05
F-statistic: 52.441
p-value: 0.0000

Decision: Reject H0
Conclusion: Sand content has a statistically significant effect on mean
compression resistance.
Interpretation: At least one mean compression strength differs among sand levels.
=====

```

Q2a. First, the test for equal variance (Levene Test) state the hypothesis

```

=====
Levene Test for Equality of Variances
=====
H0:  $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2$  (All population variances are equal)
Ha: At least one  $\sigma_i^2$  differs (Variances are not equal)
Significance Level (α) = 0.05

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  4  1.3657 0.2811
      20

=====
Levene Test Interpretation
=====
p-value = 0.2811
Decision: Fail to reject H0 → Variances are equal.
Conclusion: The assumption of equal variances is satisfied.
=====

```

2b. Provide a snip of the R results of the Levene test:

```

=====
Part 2, Question 2: Levene Test Results
=====
Objective: Test whether group variances are equal (homogeneity of variances)
Center = median

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  4  1.3657 0.2811
      20

=====
Notes for Screenshot:
- Include the section that begins with:
  'Levene's Test for Homogeneity of Variance (center = median)'
- Show the Df, F value, and Pr(>F) in your screenshot.
=====

```

2c. Interpret the results and provide a conclusion of equal variance

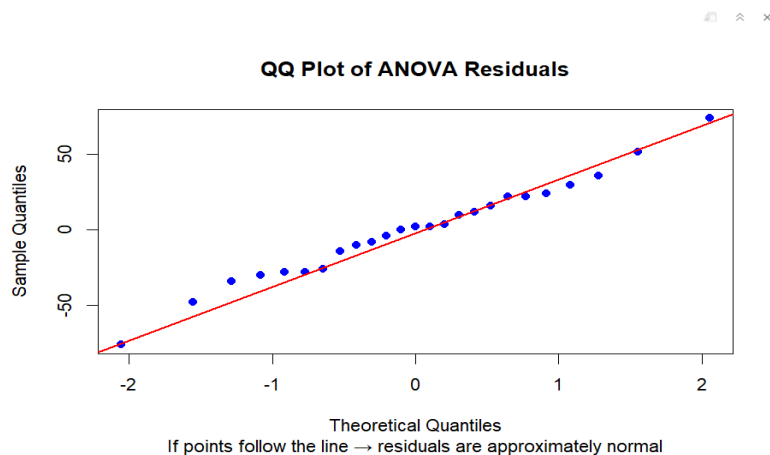
```

=====
Part 2, Question 3: Interpretation of Levene Test
=====
F-statistic = 1.366
p-value = 0.2811
Significance Level ( $\alpha$ ) = 0.05

Decision: Fail to reject  $H_0$ .
Conclusion: Variances are equal among the sand levels.
→ The equal variance (homogeneity) assumption for ANOVA is satisfied.
=====

```

2d. To test for normality, provide a qqplot of the residuals and state what you see



2e. Test for normality (Shapiro Test), state the hypothesis

```

=====
Part 2, Question 5: Shapiro-Wilk Normality Test
=====
HO: Residuals are normally distributed ( $\epsilon_i \sim N(0, \sigma^2)$ )
Ha: Residuals are NOT normally distributed
Significance Level ( $\alpha$ ) = 0.05

```

```

      Shapiro-wilk normality test

data: residuals_anova
W = 0.98894, p-value = 0.9921

```

2f. Provide a snip of the results from the Shapiro test using R

```

=====
Part 2, Question 6: Shapiro-Wilk Test Results
=====

      Shapiro-wilk normality test

data: residuals_anova
W = 0.98894, p-value = 0.9921

=====
NOTE:
- Take a screenshot of the output lines below that start with:
  'Shapiro-wilk normality test'
- Include the W statistic and p-value in your snip.
=====

```

2g. Interpret the results and provide a conclusion of normality

```

=====
Shapiro-wilk Normality Test (Residuals)
=====

      Shapiro-wilk normality test

data: residuals_anova
W = 0.98894, p-value = 0.9921

Interpretation:
p-value = 0.9921  $\geq$  0.05  $\rightarrow$  Fail to reject  $H_0$ .
Conclusion: Residuals are approximately normal.
Implication: The normality assumption for ANOVA is satisfied.
=====

```

2h. State your final Conclusion of your Overall ANOVA Test

```

=====
Part 2, Question 8: Final Conclusion of Overall ANOVA Test
=====
Significance Level ( $\alpha$ ): 0.05
Levene Test p-value: 0.2811
Shapiro-wilk Test p-value: 0.9921
ANOVA F-Statistic: 52.441
ANOVA p-value: 0.0000

✅ All assumptions are satisfied (equal variances and normality confirmed).
✅ ANOVA test is significant ( $p < 0.05$ ).
Conclusion: Reject  $H_0 \rightarrow$  Sand content significantly affects mean compression
resistance.
At least one sand percentage produces a different mean compression resistance.
ANOVA results are valid and reliable.
=====

```